LASSO for structural break estimation in time series

Chun Yip Yau

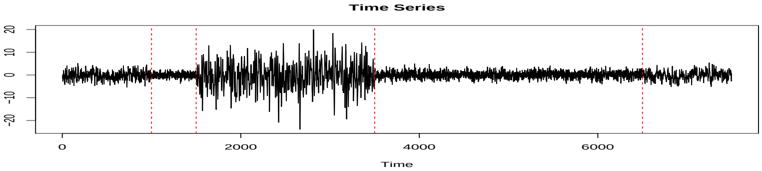
Department of Statistics
The Chinese University of Hong Kong
Research supported in part by HKSAR-RGC-GRF

27 March 2012

Joint with N.H. Chan (CUHK) and R.M. Zhang (Zhejiang University)

Content

Piecewise Stationary Time Series



- Interpreted as stationary time series with structural changes at $\{t_1, \ldots, t_m\}$.
- An intuitive model for non-stationary time series.
- *Difficulty in estimation*: The optimization

$$\arg \min_{\{t_1, t_2, \ldots, t_m\}} \sum_{i=1}^{m} L(t_i, t_{i+1}),$$

requires $\binom{n}{m}$ evaluations of $L(t_i, t_{i+1})$, the criterion function for the i-th segment $\{y_{t_i+1}, \ldots, y_{t_{i+1}}\}$.

Estimation of Piecewise Stationary Time Series

Literatures:

- Ombao, Raz, Von Sachs and Malow (2001): SLEX transformation (a family of orthogonal transformation) for segmentation.
- Davis, Lee and Rodriguez-Yam (2006,2008): Minimum Description Length (MDL) criterion function and Genetic algorithm for the optimization

$$\arg \min_{\{t_1,t_2,\ldots,t_m\}} \sum_{i=1}^{m} MDL(t_i, t_{i+1}) \,.$$

- Bayesian appraoches: (Lavielle (1998), Punskaya, Andrieu, Doucet and Fitzgerald (2002)).
- Some drawbacks:
  - computationally intensive
  - lack of theoretical justifications

The Structural Break Autoregressive (SBAR) Model

The SBAR model

$$Y_t = \begin{cases} \beta_{1,1}Y_{t-1} + \beta_{1,2}Y_{t-2} + \ldots + \beta_{1,p}Y_{t-p} + \sigma_1\epsilon_t\,, & \text{if } 1 \le t < \tau_1\,, \\ \beta_{2,1}Y_{t-1} + \beta_{2,2}Y_{t-2} + \ldots + \beta_{2,p}Y_{t-p} + \sigma_2\epsilon_t\,, & \text{if } \tau_1 \le t < \tau_2\,, \\ \ldots\ldots & \\ \beta_{m+1,1}Y_{t-1} + \ldots + \beta_{m+1,p}Y_{t-p} + \sigma_{m+1}\epsilon_t\,, & \text{if } \tau_m \le t < n\,, \end{cases}$$

can be reformulated as a high-dimensional regression framework

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} \mathbf{Y}_0^{\mathrm{T}} & 0 & 0 & \ldots & 0 \\ \mathbf{Y}_1^{\mathrm{T}} & \mathbf{Y}_1^{\mathrm{T}} & 0 & \ldots & 0 \\ \mathbf{Y}_2^{\mathrm{T}} & \mathbf{Y}_2^{\mathrm{T}} & \mathbf{Y}_2^{\mathrm{T}} & \ldots & 0 \\ \vdots & & & & \\ \mathbf{Y}_{n-1}^{\mathrm{T}} & \mathbf{Y}_{n-1}^{\mathrm{T}} & \mathbf{Y}_{n-1}^{\mathrm{T}} & \ldots & \mathbf{Y}_{n-1}^{\mathrm{T}} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \mathbf{0} \\ \boldsymbol{\beta}_2 - \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_{m+1} - \boldsymbol{\beta}_m \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \sigma_1\epsilon_1 \\ \vdots \\ \sigma_2\epsilon_{\tau_1} \\ \vdots \\ \sigma_{k+1}\epsilon_{\tau_k} \\ \vdots \\ \sigma_{m+1}\epsilon_n \end{pmatrix},$$

where

- $\mathbf{Y}_{t-1}^{\mathrm{T}} = (Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p})\,,$
- $\boldsymbol{\beta}_k^{\mathrm{T}} = (\beta_{k,1}, \ldots, \beta_{k,p})\,.$

An n-dimensional Regression Problem under Sparsity

Write

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} \mathbf{Y}_0^{\mathrm{T}} & 0 & 0 & \dots & 0 \\ \mathbf{Y}_1^{\mathrm{T}} & \mathbf{Y}_1^{\mathrm{T}} & 0 & \dots & 0 \\ \mathbf{Y}_2^{\mathrm{T}} & \mathbf{Y}_2^{\mathrm{T}} & \mathbf{Y}_2^{\mathrm{T}} & \dots & 0 \\ \vdots & & & & \\ \mathbf{Y}_{n-1}^{\mathrm{T}} & \mathbf{Y}_{n-1}^{\mathrm{T}} & \mathbf{Y}_{n-1}^{\mathrm{T}} & \dots & \mathbf{Y}_{n-1}^{\mathrm{T}} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \mathbf{0} \\ \boldsymbol{\beta}_2 - \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_{m+1} - \boldsymbol{\beta}_m \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \sigma_1 \epsilon_1 \\ \vdots \\ \sigma_2 \epsilon_{\tau_1} \\ \vdots \\ \sigma_{k+1} \epsilon_{\tau_k} \\ \vdots \\ \sigma_{m+1} \epsilon_n \end{pmatrix}$$

as

$$\vec{\mathbf{Y}}_n = \mathbf{X}_n \boldsymbol{\theta}_n + \mathbf{e}_n \, .$$

- Goal: Want a sparse solution for $\boldsymbol{\theta}_n$.
- The non-zero entries of $\boldsymbol{\theta}_n$ comprise the change-points.
  - Location estimate: $\mathcal{A}_n := \{\hat{t}_1, \dots, \hat{t}_{\hat{m}}\} = \{t : \theta_t \in \boldsymbol{\theta}_n, \theta_t \neq 0\}$.
  - Parameter estimate: $\hat{\boldsymbol{\beta}}_k = \sum_{j=1}^{\hat{t}_k} \theta_j$.

LASSO: sparse solution for regression problems

- Goal: Want a sparse solution for $\boldsymbol{\theta}_n$ for

$$\vec{\mathbf{Y}}_n = \mathbf{X}_n \boldsymbol{\theta}_n + \mathbf{e}_n \, .$$

- LASSO perfectly suits this problem:
    - obtain a sparse solution in a computationally efficient way.
- LASSO:

$$\arg\min_{\boldsymbol{\theta}_n} \frac{1}{n} \left\| \vec{\mathbf{Y}}_n - \mathbf{X}_n \boldsymbol{\theta}_n \right\|^2 + \lambda_n \sum_{i=1}^{n} \|\theta_i\| \, ,$$

where $\boldsymbol{\theta}_n = (\theta_1, \ldots, \theta_n)$, $\theta_k \in \mathcal{R}^p$.
- Tibshirani (1996): LASSO $\longrightarrow \theta_i \in \Re$
- Yuan and Lin (2005): Group LASSO $\longrightarrow \theta_i \in \Re^p$
- The challenge: dependent data.

Assumptions

The true model

$$Y_t = \begin{cases} \beta^0_{1,1} Y_{t-1} + \beta^0_{1,2} Y_{t-2} + \ldots + \beta^0_{1,p} Y_{t-p} + \sigma^0_1 \epsilon_t \,, & \text{if } 1 \le t < \tau^0_1 \,, \\ \beta^0_{2,1} Y_{t-1} + \beta^0_{2,2} Y_{t-2} + \ldots + \beta^0_{2,p} Y_{t-p} + \sigma^0_2 \epsilon_t \,, & \text{if } \tau^0_1 \le t < \tau^0_2 \,, \\ \ldots\ldots & \\ \beta^0_{m_0+1,1} Y_{t-1} + \beta^0_{m_0+1,2} Y_{t-2} + \ldots + \beta^0_{m_0+1,p} Y_{t-p} + \sigma^0_{m_0+1} \epsilon_t \,, & \text{if } \tau^0_{m_0} \le t < n \,, \end{cases}$$

LASSO:

$$\arg\min_{\boldsymbol{\theta}_n} \frac{1}{n} \left\| \vec{\mathbf{Y}}_n - \mathbf{X}_n \boldsymbol{\theta}_n \right\|^2 + \lambda_n \sum_{i=1}^{n} \|\theta_i\| \,.$$

Assumptions:

- H1: $\{\varepsilon_t\}$ i.i.d$(0,1)$ and $\mathrm{E}|\varepsilon_1|^{4+\delta} < \infty$ for some $\delta > 0$.
- H2: All characteristic roots of the AR polynomials are outside the unit circle and $\min_{1 \le i \le m_0+1} ||\boldsymbol{\beta}^0_i - \boldsymbol{\beta}^0_{i-1}|| > 0$.
- H3: $\min_{1 \le i \le m_0+1} |\tau^0_i - \tau^0_{i-1}|/(n\gamma_n) \to \infty$ for some $\gamma_n \to 0$ with $n^2 (n\gamma_n)^{-2-\delta/2} \to 0$ and $\gamma_n/\lambda_n \to \infty$.

Theorem 1

Consistency of the change-point estimates when the number of change-points is known. Assume H1, H2 and H3, and assume that $|\mathcal{A}_n| = m_0$ *is fixed in advance. If* $\lambda_n = 6pC\sqrt{\log n/n}$ *for some* $C > 1 + \sqrt{1+2b}$, $b = 2(\max_t \mathrm{E}Y_t^2 + 1)$, *then*

$$P\{\max_{1 \le i \le m_0} |\hat{t}_i - t_i^0| \le n\gamma_n\} \to 1, \quad \text{as } n \to \infty,$$

where $\gamma_n \to 0$ *with* $n^2(n\gamma_n)^{-2-\delta/2} \to 0$ *and* $\gamma_n/\lambda_n \to \infty$.

Remarks:

1. It is not possible to estimate t_i^0 consistently.
2. γ_n is interpreted as the convergence rate for the relative change-point location $\xi_i^0 = t_i^0/n$.
3. If $\mathrm{E}|\varepsilon_1|^q < \infty$ for all $q > 0$, then $\gamma_n = O(\frac{\log n}{n})$.

When the number of change-points is unknown,

- the number of change-points will not be underestimated.
- for each true change-point τ_k, there exists an estimated change-point around its $n\gamma_n$ neighborhood.

Theorem 2

If H1, H2 and H3 holds, then as $n \to \infty$,

$$P\{|\mathcal{A}_n| \geq m_0\} \to 1\,,$$

and

$$P\{\max_{b\in\mathcal{A}} \min_{a\in\mathcal{A}_n} |b-a| \leq n\gamma_n\} \to 1,$$

where $\gamma_n \to 0$ with $n^2(n\gamma_n)^{-2-\delta/2} \to 0$ and $\gamma_n/\lambda_n \to \infty$, $\mathcal{A}$ is the set of true change-points, $\mathcal{A}_n$ is the set of change-point estimates.

Two-step Estimation Procedure

- After applying LASSO, the true change-points are identified in a $n\gamma_n$ neighborhood, but the number of change-points may be overestimated, i.e. $|\mathcal{A}_n| > m_0$.
- It is natural to choose the best possible subset of $\mathcal{A}_n$ as the estimated change-points, using an information criterion of the form

$$IC(m, \mathbf{t}) = \sum_{j=1}^{m+1} \sum_{t=t_{j-1}}^{t_j - 1} (Y_t - \widehat{\widehat{\boldsymbol{\beta}}}_j \mathbf{Y}_{t-1})^2 + m\omega_n \,,$$

which is a sum of a goodness of fit measure and a penalty term, where $\widehat{\widehat{\boldsymbol{\beta}}}_j$ is the least squares estimator for the segment $\{t_{j-1}, \ldots, t_j - 1\}$.

Two-step estimation procedure

Information criterion:

$$IC(m, \mathbf{t}) = \sum_{j=1}^{m+1} \sum_{t=t_{j-1}}^{t_j - 1} (Y_t - \hat{\hat{\boldsymbol{\beta}}}_j \mathbf{Y}_{t-1})^2 + m\omega_n \, .$$

- Using the change-point estimate $\mathcal{A}_n$ from LASSO, we estimate the number and locations of the change points by

$$(\hat{\hat{m}}, \hat{\hat{\mathbf{t}}}) = \arg \min_{\substack{m \in (0,1,\dots,|\mathcal{A}_n|), \\ \mathbf{t}=(t_1,\dots,t_m) \subset \mathcal{A}_n}} IC(m, \mathbf{t}) \, .$$

- Examples:
  - BIC of Yao (1988)
  - MDL of Davis, Lee and Rodgriduez-Yam (2006,2008)
- Computational burden reduces from $\binom{n}{m}$ to $2^{|\mathcal{A}_n|}$.

Consistency of the change-point locations when the number of change-points is unknown:

$$(\hat{\hat{m}}, \hat{\hat{\mathbf{t}}}) = \arg \min_{\substack{m \in (0,1,\ldots,|\mathcal{A}_n|), \\ \mathbf{t}=(t_1,\ldots,t_m) \subset \mathcal{A}_n}} IC(m, \mathbf{t}) \, .$$

Theorem 3

Assume that H1, H2 and H3 hold and assume that the penalty term ω_n *satisfies* $\lim_{n\to\infty} \omega_n / [8 m_0 n \gamma_n (\max_{1\le i \le n} \mathrm{E} Y_i^2)] > 1$. *Further assume that* $t_i^0 = [n\xi_i^0]$ *with* $\min_{1\le i\le m_0} |\xi_i^0 - \xi_{i-1}^0| \geq \varepsilon > 0$. *Then*

$$P\{\widehat{\widehat{m}} = m_0\} \ \rightarrow \ 1 \, ,$$

and

$$P\{\max_{1\le i\le m_0} |\widehat{\widehat{t}}_i - t_i^0| \leq n\gamma_n\} \ \rightarrow \ 1.$$

Two-step Estimation Procedure

When the number of change-points is unknown, the estimator is

$$(\hat{\hat{m}}, \hat{\hat{\mathbf{t}}}) = \arg \min_{\substack{m \in (0,1,\ldots,|\mathcal{A}_n|), \\ \mathbf{t}=(t_1,\ldots,t_m) \subset \mathcal{A}_n}} IC(m, \mathbf{t}) \,.$$

- It requires $2^{|\mathcal{A}_n|}$ evaluations of the IC.
- If $2^{|\mathcal{A}_n|}$ is too large we can further simplify the computation by the *backward elimination algorithm* (BEA).
- BEA further reduces the computational order from $2^{|\mathcal{A}_n|}$ to $|\mathcal{A}_n|^2$.

Backward Elimination Algorithm (BEA)

- The BEA starts with the set of change-points $\mathcal{A}_n$, then
- removes the "most redundant" change-point that corresponds to the largest reduction of the IC.
- repeat successively until no further removal is possible.

1. Set $K = |\mathcal{A}_n|$, $\mathbf{t}_K := (t_{K,1}, \dots, t_{K,K}) = \mathcal{A}_n$ and $V_K^* = IC(K, \mathcal{A}_n)$.
2. For $i = 1, \dots, K$, compute $V_{K,i} = IC(K-1, \mathbf{t}_K \backslash \{t_{K,i}\})$. Set $V_{K-1}^* = \min_i V_{K,i}$.
3.
    - If $V_{K-1}^* > V_K^*$, then the estimated locations of change-points are $\mathcal{A}_n^* = \mathbf{t}_K$.
    - If $V_{K-1}^* \leq V_K^*$ and $K = 1$, then $\mathcal{A}_n^* = \emptyset$. That is, there is no change-point in the time series.
    - If $V_{K-1}^* \leq V_K^*$ and $K > 1$, then set $j = \arg\min_i V_{K,i}$, $\mathbf{t}_{K-1} := \mathbf{t}_K \backslash \{t_{K-1,j}\}$ and $K = K - 1$. Go to step 2.

Backward Elimination Algorithm (BEA)

Example

- LASSO gives the estimate $\mathcal{A}_n = (\hat{t}_1, \hat{t}_2, \hat{t}_3)$.
  - $V_3^* = IC(3, \mathcal{A}_n) = 10$.

1. Removing one point:
   i. $V_{3,1} = IC(2, (\hat{t}_1, \hat{t}_2)) = 11$.
   ii. $V_{3,2} = IC(2, (\hat{t}_1, \hat{t}_3)) = 10.5$.
   iii. $V_{3,3} = IC(2, (\hat{t}_2, \hat{t}_3)) = 9$.
   - $V_2^* = \min_i V_{3,i} = 9 <= V_3^* = 10$, proceed for further reduction.
2. Removing one more point:
   i. $V_{2,1} = IC(1, (\hat{t}_2)) = 10$
   ii. $V_{2,2} = IC(1, (\hat{t}_3)) = 9.5$.
   - $V_1^* = \min_i V_{2,i} = 9.5 > V_2^* = 9$

- Conclude that $\hat{\hat{m}} = 2$, $\hat{\hat{\mathbf{t}}} = (\hat{t}_2, \hat{t}_3)$.

Consistency of the change-point estimates when the number of change-points is unknown:

Theorem 4

Let $\mathcal{A}_n^* =: (\hat{t}_i^*, \ldots, \hat{t}_{|\mathcal{A}_n^*|}^*)$ *be the estimate obtained from BEA. Under the conditions of Theorem 3, we have*

$$P\{|\mathcal{A}_n^*| = m_0\} \;\to\; 1\,,$$

and

$$P\{\max_{1\le i\le m_0} |\hat{t}_i^* - t_i^0| \le n\gamma_n\} \;\to\; 1.$$

Summary: Two-step procedure of change-point estimation

- First Step: Get a possibly overestimated locations estimator $\mathcal{A}_n$ from the LASSO

$$\arg\min_{\boldsymbol{\theta}_n} \frac{1}{n} \left\| \vec{\mathbf{Y}}_n - \mathbf{X}_n \boldsymbol{\theta}_n \right\|^2 + \lambda_n \sum_{i=1}^{n} \|\theta_i\| ,$$

- Second Step: Select the best subset of change-points from $\mathcal{A}_n$ by the Information Criterion

$$(\hat{\hat{m}}, \hat{\hat{\mathbf{t}}}) = \arg \min_{\substack{m \in (0,1,\ldots,|\mathcal{A}_n|), \\ \mathbf{t}=(t_1,\ldots,t_m) \subset \mathcal{A}_n}} IC(m, \mathbf{t}) .$$

  - If $2^{|\mathcal{A}_n|}$ is not large, then all possible subsets can be evaluated.
  - Otherwise, Backward Elimination Algorithm can be used to obtain the location estimates.

- Consistency:

$$P\{|\mathcal{A}_n^*| = m_0\} \rightarrow 1 ,$$

$$\text{and} \quad P\{\max_{1 \le i \le m_0} |\hat{t}_i^* - t_i^0| \le n\gamma_n\} \rightarrow 1.$$

Computational Issues of Group LASSO

Two fast implementations of group LASSO in the first step:

$$\arg\min_{\boldsymbol{\theta}_n} \frac{1}{n} \left\| \vec{\mathbf{Y}}_n - \mathbf{X}_n \boldsymbol{\theta}_n \right\|^2 + \lambda_n \sum_{i=1}^{n} \|\theta_i\| .$$

1. Exact Solution by block coordinate descent. (Yuan & Lin (2006), Fu (1998)):
   - iteratively solving estimating equations
   - converges to the global optimum
   - stable and efficient
2. Approximate Solution by group Least Angle Regression (LARS). (Erfon et al. (2004), Yuan & Lin (2006)):
   - add the "most correlated" covariate one by one.
   - computationally more efficient.
   - well approximates the solution of group LASSO in many cases.

- When $p = 1$, LARS algorithm gives the exact solution of LASSO.

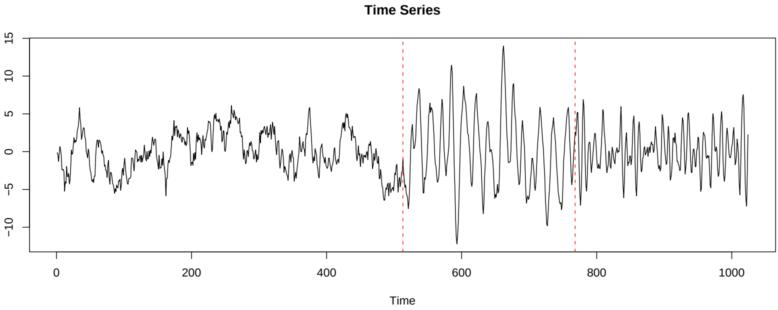
Model Selection in each segment

- A stationary AR(p) model is assumed in each segment.
- The theoretical results hold if p is greater than the maximum order among all segments.
- In practice, a large p (e.g., $p = 10$) is used in the two-step estimation procedure.
- After the change-points are detected, standard model selection procedure can be applied for each segment.
- Since the convergence rate of change-point locations is faster than $n^{-1/2}$, the model selection has the same asymptotic properties as the no-change-point case.

Example 1. Compare to Davis, Lee and Rodgriduez-Yam (2006)

- True model:

$$Y_t = \begin{cases} 0.9Y_{t-1} + \epsilon_t\,, & \text{if } 1 \le t \le 512\,, \\ 1.69Y_{t-1} - 0.81Y_{t-2} + \epsilon_t\,, & \text{if } 513 \le t \le 768\,, \\ 1.32Y_{t-1} - 0.81Y_{t-2} + \epsilon_t\,, & \text{if } 769 \le t \le 1024\,. \end{cases}$$



Example 1. Compare to Davis, Lee and Rodgriduez-Yam (2006)

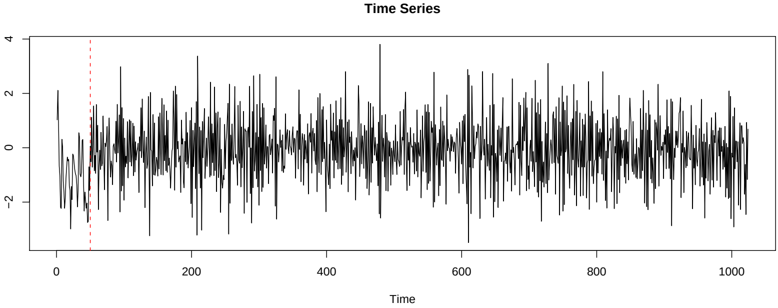
- True relative location of change-points $= (\frac{512}{1024}, \frac{768}{1024}) = (0.5, 0.75)$.
- Replications: 200.

Number of	Auto-PARM			Two-Step		
segments	(%)	Mean	SE	(%)	Mean	SE
3	96.0	0.500	0.007	100	0.500	0.012
		0.750	0.005		0.750	0.011
4	4.0	0.496	0.004	0		
		0.566	0.108			
		0.752	0.003			

Example 2. Compare to Davis, Lee and Rodgriduez-Yam (2006)

- True model:

$$Y_t = \begin{cases} 0.75Y_{t-1} + \epsilon_t\,, & \text{if } 1 \le t \le 50\,, \\ -0.5Y_{t-1} + \epsilon_t\,, & \text{if } 51 \le t \le 1024\,. \end{cases}$$



Example 2. Compare to Davis, Lee and Rodgriduez-Yam (2006)

- True relative location of change-points $= (\frac{50}{1024}) = (0.0488)$.
- Replications: 200.

Number of segments	Auto-PARM (%)	Auto-PARM Mean	Auto-PARM SE	Two-Step (%)	Two-Step Mean	Two-Step SE
2	100	0.042	0.004	100	0.049	0.004

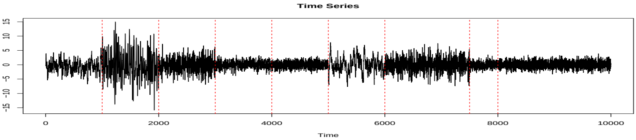
Example 3. Long time series with many change-points

- True model:

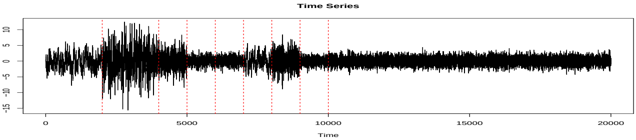
$$Y_t = \begin{cases} 0.9Y_{t-1} + \epsilon_t, & \text{if } 1 \leq t \leq t_1, \\ 1.69Y_{t-1} - 0.81Y_{t-2} + \epsilon_t, & \text{if } t_1 \leq t \leq t_2, \\ 1.32Y_{t-1} - 0.81Y_{t-2} + \epsilon_t, & \text{if } t_2 \leq t \leq t_3, \\ 0.7Y_{t-1} - 0.2Y_{t-2} + \epsilon_t, & \text{if } t_3 \leq t \leq t_4, \\ 0.1Y_{t-1} - 0.3Y_{t-2} + \epsilon_t, & \text{if } t_4 \leq t \leq t_5, \\ 0.9Y_{t-1} + \epsilon_t, & \text{if } t_5 \leq t \leq t_6, \\ 1.32Y_{t-1} - 0.81Y_{t-2} + \epsilon_t, & \text{if } t_6 \leq t \leq t_7, \\ 0.25Y_{t-1} + \epsilon_t, & \text{if } t_7 \leq t \leq t_8, \\ -0.5Y_{t-1} + 0.1Y_{t-2} + \epsilon_t, & \text{if } t_8 \leq t \leq T. \end{cases}$$

- Three scenarios of relative change-point locations.
  - $\boldsymbol{\tau}$ =(0.1,0.2,0.3,0.4,0.5,0.6,0.75,0.8)
  - $\boldsymbol{\tau}$ =(0.1,0.2,0.25,0.3,0.35,0.4,0.45,0.5)
  - $\boldsymbol{\tau}$ =(0.1,0.2,0.25,0.3,0.5,0.8,0.9,0.95)

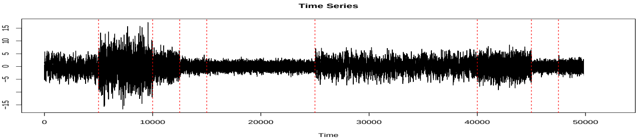
- Scenario 1: $\boldsymbol{\tau} = (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.75, 0.8); n = 10,000.$



- Scenario 2: $\boldsymbol{\tau} = (0.1, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5); n = 20,000.$



- Scenario 3: $\boldsymbol{\tau} = (0.1, 0.2, 0.25, 0.3, 0.5, 0.8, 0.9, 0.95); n = 50,000.$

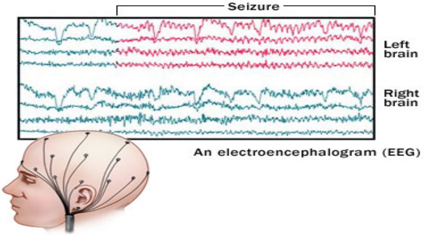


Results from two-step estimation procedure. Replication=200.

	Scenario 1			Scenario 2			Scenario 3		
T	10000			20000			50000		
Computing Time	4s			7s			18s		
% of $\hat{m} = 8$	90			84			92		
	True	Mean	SE	True	Mean	SE	True	Mean	SE
t_1/T	0.1	0.1022	0.0091	0.1	0.1001	0.0010	0.1	0.1020	0.0129
t_2/T	0.2	0.2008	0.0012	0.2	0.1998	0.00042	0.2	0.1999	0.00018
t_3/T	0.3	0.3001	0.0010	0.25	0.2499	0.00048	0.25	0.2500	0.00020
t_4/T	0.4	0.3942	0.0088	0.3	0.2984	0.0032	0.3	0.2998	0.00039
t_5/T	0.5	0.4999	0.0012	0.35	0.3501	0.00090	0.5	0.4999	0.00020
t_6/T	0.6	0.5999	0.0010	0.4	0.4001	0.00081	0.8	0.7999	0.00026
t_7/T	0.75	0.7501	0.0011	0.45	0.4501	0.00057	0.9	0.9000	0.00021
t_8/T	0.8	0.7998	0.0016	0.5	0.4998	0.00070	0.95	0.9499	0.00044

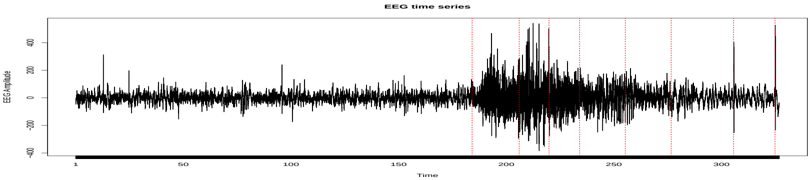
Electroencephalogram (EEG) Time Series

- EEG is the recording of electrical activity along the scalps of a subject.
- It measures voltage fluctuations resulting from ionic current flows within the neurons of the brain.
- use for medical diagnostics:
    - epilepsy
    - tumors
    - stroke
    - brain disorders ...



Electroencephalogram (EEG) Time Series

- The following EEG is recorded from a female patient diagnosed with left temporal lobe epilepsy.
- Data collection:
  - Sampling rate: 100Hz,
  - Recording period: 5 minutes and 28 seconds,
  - Sample size: n=32,768.
- Investigated by Ombao *et al*. (2000) and Davis, Lee and Rodriguez-Yam (2006).
- Results of Two-step LASSO procedure:



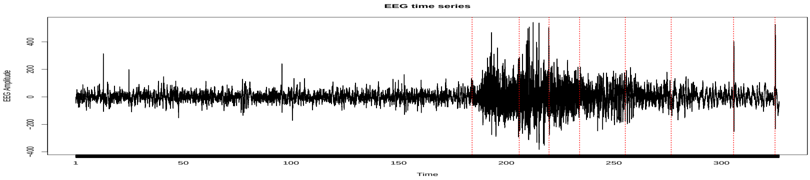
Electroencephalogram (EEG) Time Series

	Locations of change points (seconds)										
	1	2	3	4	5	6	7	8	9	10	11
Two-step	184.2	206.1	220.0	234.2	255.4	276.7	305.7	325.0	-	-	-
Auto-PARM	185.8	189.6	206.2	220.9	233.0	249.0	261.6	274.6	306.0	308.4	325.8

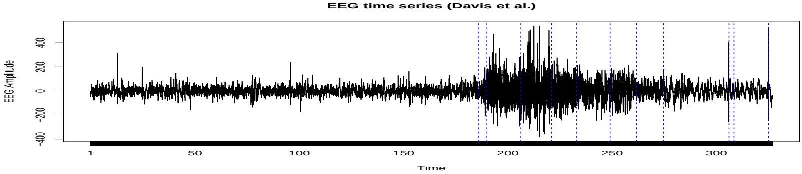
- Estimated starting time for seizure:
  - Neurologist: t=185s.
  - Auto-PARM: t=185.8s.
  - LASSO Two-step: t=184.23s.

Electroencephalogram (EEG) Time Series

- Two-step Lasso procedure

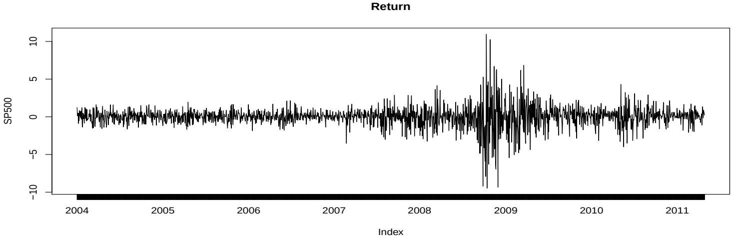


- MDL+Genetic algorithm (Davis, Lee and Rodriguez-Yam (2005))



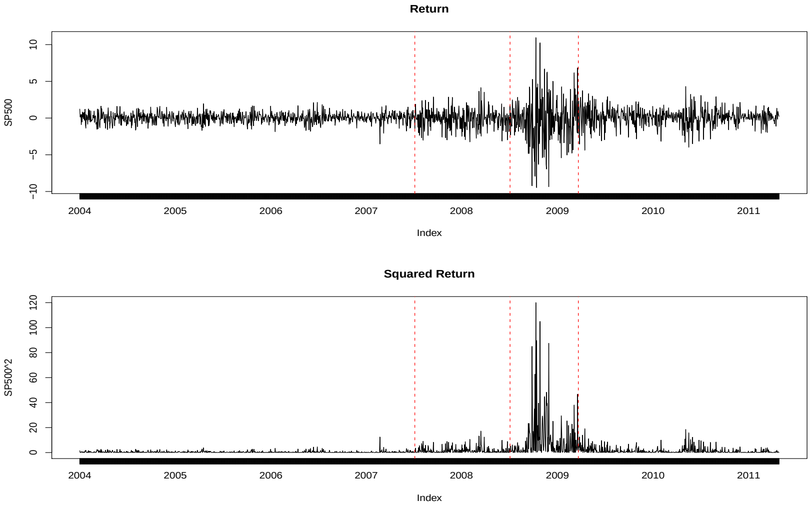
Standard & Poor's 500 Index

- From Jan 2, 2004 to April 29, 2011.
- Structural changes in Volatility.

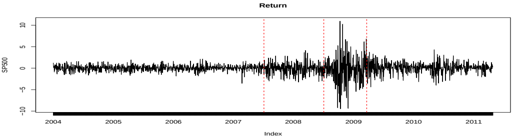


- Applications of LASSO procedure for structural changes in volatility.
  - Two-step LASSO procedure assumes piecewise stationary time series with an autoregressive structure.
  - Volatility is modeled by GARCH processes.
  - The square of the GARCH process is an ARMA process, which can in turn be approximated by AR processes.
  - Thus, the LASSO procedure can be applied to the squares of the log-return process of S&P500 series for change-point detection.

Results:


Return
SP500
10
5
0
−5
−10
2004
2005
2006
2007
2008
2009
2010
2011
Index
Squared Return
SP500^2
120
100
80
60
40
20
0
2004
2005
2006
2007
2008
2009
2010
2011
Index

Standard & Poor's 500 Index



Interpretations of the three estimated change-points:

- July 10, 2007
  - Standard and Poor's placed 612 securities backed by subprime residential mortgages on a credit watch preluded the panic of the market.
- September 15, 2008
  - Lehman Brothers Holdings incorporated filed for bankruptcy protection triggered the financial crisis.
- April 7, 2009
  - Quantitative Easing (QE) policy.
  - US Federal Reserve gradually purchased around $ 1 trillion debt, Mortgage-backed securities and Treasury notes in the early 2009 stabilized the market.

Conclusion

- Two step change point estimation procedure.
  - First step: LASSO for screening out the change-points.
  - Second step: Change-point estimation by Information Criterion.
- Consistency is proved for:
  - the estimated number of change-points.
  - the estimated locations of change-point.
- Computational efficiency:
  - LASSO: efficient.
  - LARS approximation: highly efficient.

Thank You!